Business Agility and the Cloud

A paper on how cloud technologies increase business agility

June 2011

Microsoft

Business Agility and the Cloud

In the 1950s almost no organization owned a computer. Computing was consumed as a service rather than owned as a capital asset.

Computer companies provided services in those days because the computers themselves were too expensive for most organisations to own. But we also have to think of the risk associated with the acquisition of such an expensive resource. In the '50s, computers were new-fangled gadgets and nobody really knew if there'd be any advantage to using them: they were "scientific" devices. How could science apply to a haulage business or an insurance company?

However, their use grew. Organisations outsourced their calculation problems from teams of people who took days, weeks or months to come up with an answer, to computer companies, for whom the greatest delay was the delivery of the program and the delivery of its results..

This extra speed gave these companies *business agility*. It meant they could respond more quickly to events than their competitors and could either contain profitability or improve it and often take market share.

We find ourselves in a similar position today. Large computer companies with masses of compute resources and storage capacity are offering immensely scalable compute services on a pay-as-you-go and only-pay-for-what-you-use model . We send our computer programs to them and they run them for us.

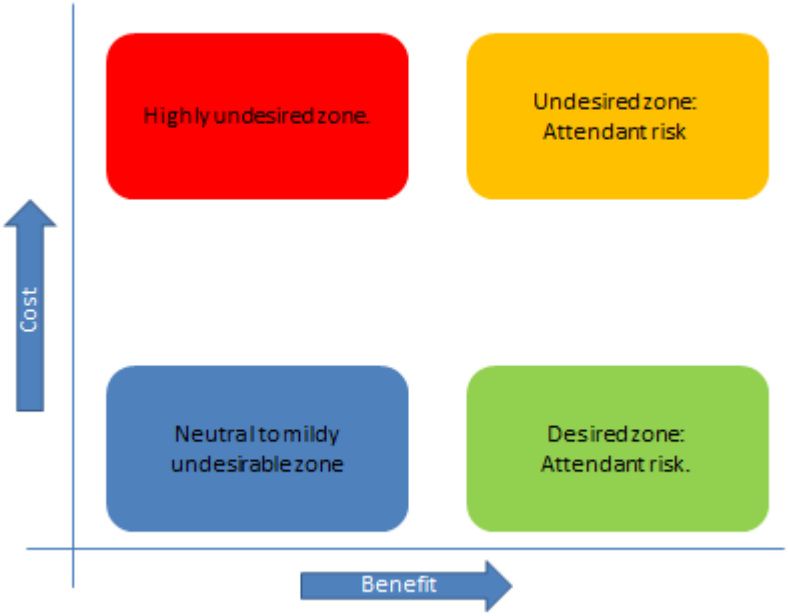
Fashion and IT make unhappy bedfellows

Over the years, most died-in-the-wool IT managers have seen new ideas come and go. They've patted themselves on the back for not getting involved in many of them. But every so often there are inflection points that are so profound they warrant more than being dismissed as *IT fashion*. The arrival of the PC and client/server computing was one big shift that started in earnest approximately 15 years ago, though it had certainly been simmering on a low heat for many years before that. The way the Internet affected the way billions of people interact with each other and with applications that serve their needs is another example. Some companies ignored these seismic shifts and while the rest of the world moved on, they lagged behind typically with greater costs, lower efficiency and poor responsiveness.

Agility and economy

Any business, as long as it has enough capital, can be more agile. When we think of some of the problems with agility, the most obvious one is that of having enough staff to take their eye off operational matters and concentrate instead on something new. With a finite set of resources (the people in the organization) to deploy, they can do one thing, or they can do another. An incredibly small percentage of them can do both – for a short time. However, a healthy bank balance allows us to either hire the people we need or buy the services we need. This may well be a good plan, but it's

a careful balancing act because eating in to the cash reserves of a company limits its potential to increase budget for research, development, exploration, business development, service development and so on. But solving some other non-operational issue might in its turn increase the profitability of the organization and therefore more than compensate for the loss of capital. There is a cost to the benefit – it's the classic Cost-Benefit Analysis problem or CoBA. In a tight-margin business, CoBA relies on accurate predicted forecasts. Though, as physicist Niels Bohr once said – "making predictions is difficult – especially about the future".



There is only one place to be on the graph above: bottom right. However, there is an attendant risk associated with projects aiming for this area. It's that the costs may be difficult to forecast and the result may push projects in to the top right hand corner. If the benefits are inaccurately forecast it merely moves the result in to the bottom-left quadrant of the graph: an undesirable place but not an unmitigated disaster. What is needed is a way to contain costs and minimize them so that any project is limited to how far it can rise on the vertical axis.

If some aspects of agility can be acquired at a cost that is easily affordable and perhaps even lower than normal operating costs for the business, this surely warrants further investigation. In their paper "The Economics of the Cloud", Rolf Harms and Michael Yamartino highlight the economies of scale that cloud computing can offer a business that consumes its IT service the traditional way: from their own datacenters. I assume in this paper that cloud services can be acquired economically and aim to highlight the next, and possibly higher-order benefit to an organization of a move to the cloud – business agility.

The 4 business profiles

The agility that cloud computing (at an affordable cost) offers can be realized by businesses that follow these profiles. Business profiles in this sense are not divided in to vertical or horizontal

sectors, but in their appetite for either consumption or production. One business may exhibit many profiles and have different profiles for consumption than they do for production.

The 4 types are:

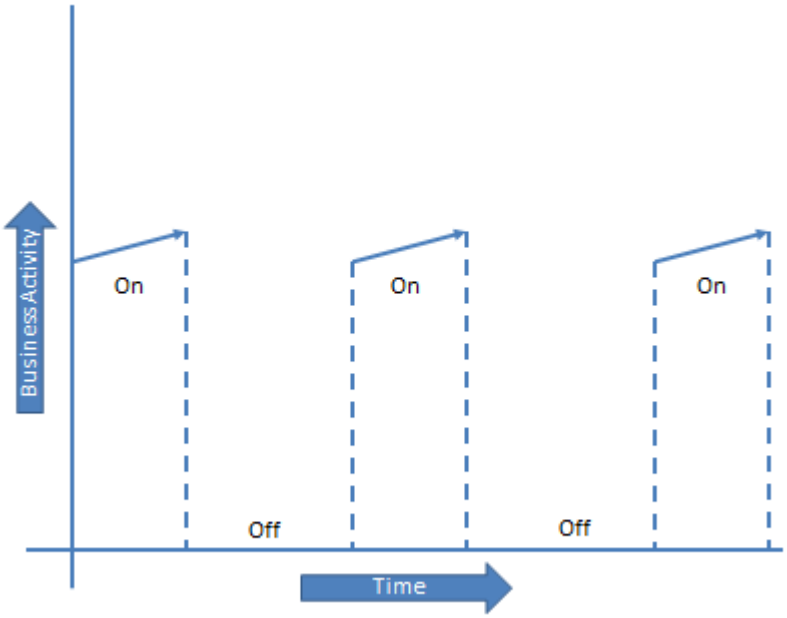
1. On-off
2. Growing fast
3. Predictable bursting
4. Unpredictable bursting

A business which experiences no need to change, is not under competitive pressures, not subject to changes in working practices, legislation, materials and so on has no need to become agile. I don't believe there are many of these around! Most businesses constantly have to evolve and adapt as the world changes around them. Competitors beat them to market or manufacture their products more cheaply. Legislation is thrown at them. Standards of reporting for compliance get tougher every year. The raw materials they use become scarcer. The services they offer go out of fashion. The media causes troughs and peaks in their business map. The practices of the workforce change over time. The shareholders randomly change their mind from year to year. These sort of changes can be seen sometimes across the entire business (or the entire industry, the entire country or even the entire world economy - think of the recent banking crisis) and almost definitely within departments of a business. These sorts of changes can be characterized in to one of the above profiles.

On-off

This is usually some kind of batch-work. For example a department has to create financial business reporting once every quarter. They are not in a position to spread this out over the entire quarter because they can't predict the business performance of the latter days of the quarter and the calculations are long, iterative and complex. They are forced to wait until the entire business performance figures for the quarter are in and only then can the calculation engine get to work.

Perhaps an insurance company has to run a risk analysis of the policies on its books every 50 days. Again, it can't start early, gathering data as it goes because data in the matrix affects other data in the matrix. All the data has to be available as one amorphous mass for the calculations to work.



There are clear economic reasons why a business may want a pay-as-you-go and only-pay-for-what-you-use model for this type of processing. The cloud offers this. But how does the cloud affect agility in this example.

Let's take the insurance company example. Imagine they use a large amount of computing resource every 50 days to perform these calculations. Somebody in the organization has already worked out they can't afford to keep that amount of computing capacity idle for 49 days out of every 50. So they devise a usage schedule that sees the computing resources used by other departments in the intervening period. Sometimes this involves the heavy-lifting of manual reconfiguration and so on. No problem though, it's all been compiled in to an operational procedures manual and reduced to a set of daily operations that all just happen like clockwork; executed by the IT department.

2 days after the risk calculations (which determines the pricing of policies sold in the future), let's imagine a large natural disaster such as an earthquake or hurricane, means the insurance company needs to do another risk calculation. But according to the daily procedures manual, it's not that department's turn to use those compute resources for another 48 days. To try to change the schedule will be disruptive to the rest of the business. To try and re-write the daily procedures manual "on the fly" will be impossible.

This company now has to either disrupt the rest of the business with unpredictable consequences (remember we said "making predictions is difficult – especially about the future"), or simply wait their turn in the queue and let 48 days pass. It's a dilemma that so many businesses are unwilling to get in to that they live with the uneconomic option of dedicating computing resources to departments that will sit idle, doing nothing for much of the time – this is an expensive way of buying agility. Or they forget agility and hope nothing ever occurs that will cause the procedures to be run out of their normal scheduled sequence.

When such a workload is moved to the cloud, not only is it more economical, it also permits greater agility. When the natural disaster occurs, it's simply a matter of either bringing that period's processing forward by 48 days and from that point on, operating on a new schedule. Or simply inserting a processing slot in to the unscheduled time period and then reverting to the normal schedule later. Because making the resources in the cloud available for such an unplanned eventuality is trivial, the company that adopts such a practice will be more agile than its competitors who still insist on doing things the traditional way. A huge calculation with new data such as I've mentioned could very probably be kicked off within 30 minutes or less of the disaster occurring.

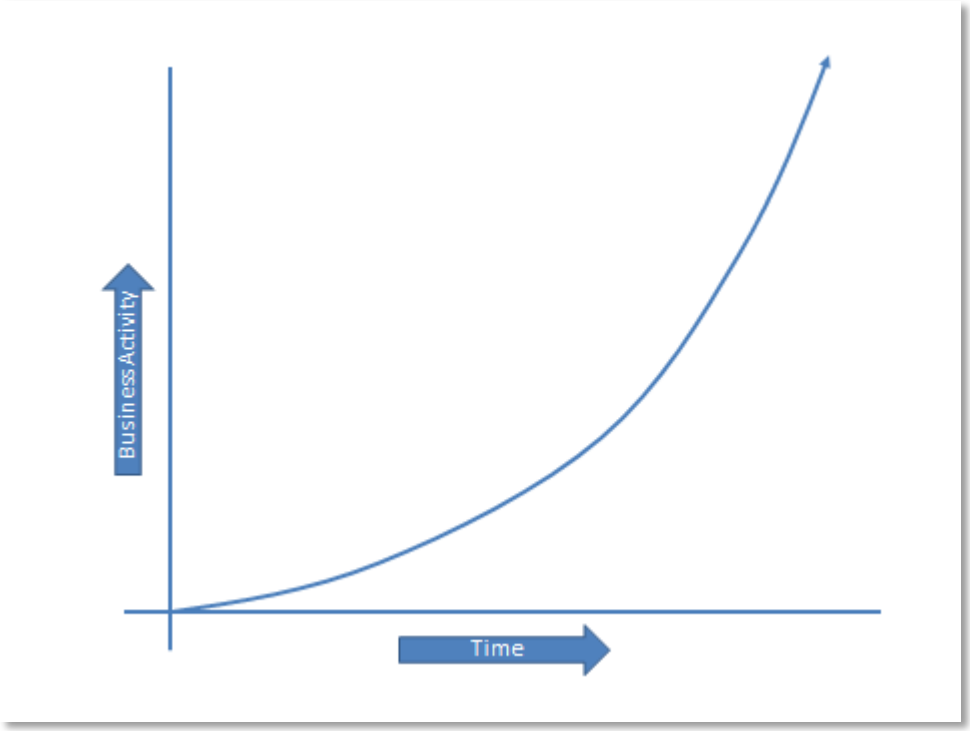
Let's continue the analogy of the natural disaster – as the story progresses, things may get better, or they may get worse. The organisation that is in a position to run such calculations let's say every 24 hours, because the resources are instantly available to them can stay on top of developments. If things get worse, they can always have available, policy prices which reflect the true risk and still allow them to make a profit. If things get better, they can sell policies more cheaply and still maintain profit margins. When they sell policies more cheaply than their competitors they take market share – the fact they can do this in the full knowledge that the policies will hit the correct profitability numbers means the sales and marketing operations need not be hamstrung by delays. The last thing a large corporate customer wants to hear is "I'll get back to you next week with a price".

Let's be clear here – the organisations involved in these kinds of workloads can of course buy and maintain a big enough datacentre to have the capacity they need to always be able to run the risk calculations. But they are using valuable capital to "buy agility" and this will negatively impact profitability. With cloud solutions, these organisations will "buy only the agility they need, when they need it": such a prospect will positively impact profitability.

Growing Fast

This workload is typically characterised by the startup company, but it could equally as well be a new sales division within an existing large company, or simply a new department.

In this workload we again meet our old friend, Niels Bohr: "making predictions is difficult – especially about the future". It's because although every new venture wants to grow fast, whether it will do or not is unpredictable, as is the rate of growth (or in some cases shrinkage!).



When a business starts, it typically has a small number of customers and over time the number grows. There are really two main issues here. One impacts profitability, the other agility.

Investing in resources before the customers have been acquired is generally seen as a prudent measure. However, over-investing in resources that are either never used, or not used for say 6 months will negatively impact profitability. Being able to quickly provision resources to cope with unpredicted customer demand is an agility problem. In say, an online business, if a new customer visits the website and has a bad experience because it is slow, unresponsive or unreliable, there is a greater loss than that single customer's business. As we all know, satisfied customers share their experiences with 1 other person but dis-satisfied customers share their experiences with 10 other people.

It is vital that the business is able to respond in this way because a product or service which catches the imagination of customers might suddenly result in sky-rocketing numbers of visitors.

The events which trigger fast growth are entirely unpredictable. Sometimes the product or service is so compelling that fast organic growth occurs naturally. Other times it can be a business that has had slow but steady growth for say 12 months, but then gets a heavy and positive mention by an influential celebrity on peak-time TV. This sort of event can result in a well scaled website reaching maximum capacity and creating large numbers of unhappy customers, each sharing their bad story with 10 others, within 15 minutes of the TV event which initiated the problem. 24 hours later, the TV might have a negative news story about how the celebrity endorsed the product but the site was unable to cope causing mass dis-satisfaction with tens or even hundreds of thousands of potential customers. The negative news story can simply exacerbate the problem halting the previous slow but steady growth and turning things in a negative direction.

The cloud increases agility in these situations because there becomes a direct relationship between the cloud resources used and the number of visitors to the site. So the cost of cloud computing resources is proportional to the business the site generates. Being able to scale the cloud resources within minutes of a significant event like this is true agility.

Again, let's be clear. A business could buy and maintain a huge datacentre at the outset and live with those costs based on the notion that the business will eventually make enough money to pay for it. What the cloud does is permit agility in a profitable way: where the cost and the benefit are directly related to each other.

It's analogous to what happened in the early days of the Cold War. Fully crewed aircraft sat on airfields "burning and turning" their engines in case a fast deployment was required. It was only as engine and aircraft technology advanced that aircraft could save on fuel and engine-hours when aircraft could be made flight-ready in similar times to their "burn and turn" predecessors.

As a general rule, cloud operators such as Microsoft have built their cloud data-centers in such a way that fast, simple, scalable and reliable provisioning of compute and storage resources is a central tenet of their operations. The technology is so advanced that most organisations' on-premise data-centers could never hope to achieve the same speed of resource provisioning.

Microsoft's Windows Azure platform doesn't simply deploy and scale-up/scale-down computing resources. It provisions entire working applications, along with all the attendant compute, storage and infrastructure services required. In many cases, massively complex applications can be provisioned in less than 10 minutes.

Let's say a new beauty-products startup business is seeing 100 visitors per day to its website. A celebrity (with no particular relationship to the business operators) on an early evening national-TV chat show endorses the product and says that is why she looks so young. Within 10 minutes of the end of the program there are 5,000 visitors and within an hour there have been 50,000 visitors.

A Windows Azure based e-commerce application is continually monitoring itself and sees the sudden peak. The monitoring causes a trigger to be fired within the application which requests more scale from Windows Azure. The site was initially running on 2 fairly low-spec servers. Within 10 minutes it is running on 8 super-high-spec servers and easily dealing with the increased demand. The success of customers placing orders, breeds more success as each one of them tells a friend of their successful experience and 15% of those friends also place orders.

With this entirely unpredictable burst in popularity, the product has become a mainstay of the beauty and fashion industry. The site now runs on 4 mid-spec servers for steady state operations but can scale within minutes to 10s or even 100s of servers whenever unpredictable growth occurs suddenly.

It's also important to realise that such scaling should be able to reverse as quickly as it can be created. Scaling back after fast growth is also an important profitability consideration. Windows Azure scales in both directions with equal ease meaning that agility is met within the bounds of profitability.

The notion of being able to bail-out of a bad business idea is a core idea behind the cloud. Some startup businesses which look great on paper are doomed to failure, but the only way to truly find out is to try the business out for real. If a business is going to fail, it's preferable that it fail without incurring un-necessary expense to the owners/investors – the notion of the so called "fail-fast, fail-cheap" idea. A large datacentre full of hardware, software and specialised equipment such as 3-phase power supplies, backup units, high-volume air-conditioning, fire and safety equipment is a massive investment. Such expense is the traditional risk a startup has to cater for. These risks don't exist with cloud applications. All the specialised equipment is owned and operated by the cloud operator and will continue to be owned and operated long after the unsuccessful startup has failed. The failed startup will only have paid for the relatively few hours (days/weeks/months) of resources it used while trying to establish itself. So cloud technologies increase agility by not only offering agility with profitability, but also offering agility with lower risk.

Predictable Bursting

The good news with this workload is that it is, by its nature, predictable. It's different to on-off workloads because there is always some form of processing the business is engaged in but there may be known, predictable times when bursts occur. Many of these bursts of activity might be self-triggered by the business. For example if an online retailer runs a monthly marketing campaign, they can usually expect a peak in web site visits for those products that are part of the campaign and of course any campaign content that supports it, such as text, images and video.

There are other times when a known event collides with a business's execution. The festive season for an online retailer. Late winter/early spring for summer-sun holiday companies. The Super-Bowl for a home-delivery pizza company. These events are known about usually many months/years in advance and are often cyclical.

It is though still the case that although the timing of the event itself is predictable, what is required of the technology infrastructure is still unpredictable. Imagine a company that prepares high volumes of cloud computing resources say 1 hour before an expected massive increase in traffic only to find that the expected peak never reached the dizzy heights predicted. Being able to shed large amounts of cloud resource quickly and simply is as important as being able to provision it quickly just before the expected event.

Trying to operate a business with this kid of flexibility and agility can stifle an organisation's internal operations when they try to do it from their owned capital assets. Acquiring the necessary hardware, software, cooling, floorspace, staff etc in advance of the expected peak dilutes the effectiveness of the organisation to carry on its core business – selling holidays, beauty products, airline tickets and so on.

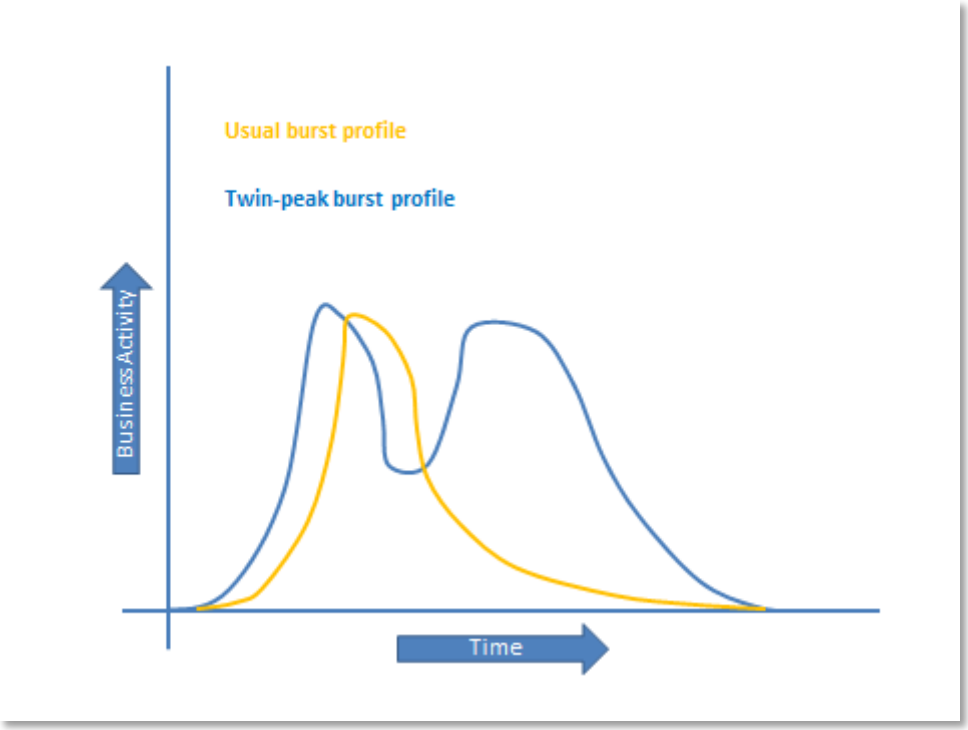
Often a firm can create some core on-premise infrastructure and build a hybrid cloud which sees most of their normal, every-day, steady state operations serviced out of the on-premise environment, but when extra capacity is required, they can "burst to the cloud".

In a business where the timing of this bursting is predictable, it allows the company to fine-tune the level of its bursting on a minute-by-minute basis. It can closely monitor the total workload in the

peak and burst more or less in to the public cloud to try and exactly match the resources to the requirements.

Any business which can take advantage of this bursting in a way that still allows for profitable operations is not only showing greater agility than their competitors but is also still making money. They are not just simply "buying" agility, in a similar way that some companies "buy" market share in the hope of edging their competitors out of a business area entirely. This is usually done at great cost and represents enormous risk to a business. Predictably bursting to the cloud, fine-tuning the balance between requirements and delivery as you go has a massive impact on the risks.

The enemy of the business with the predictable bursting profile is the *twin-peak* and the *double-dip.* The natural assumption of most organisations that live within this profile is that the burst will look roughly the same as it always has done. But life in unpredictable, and those organisations that are positioned to capitalise on these events are better placed.

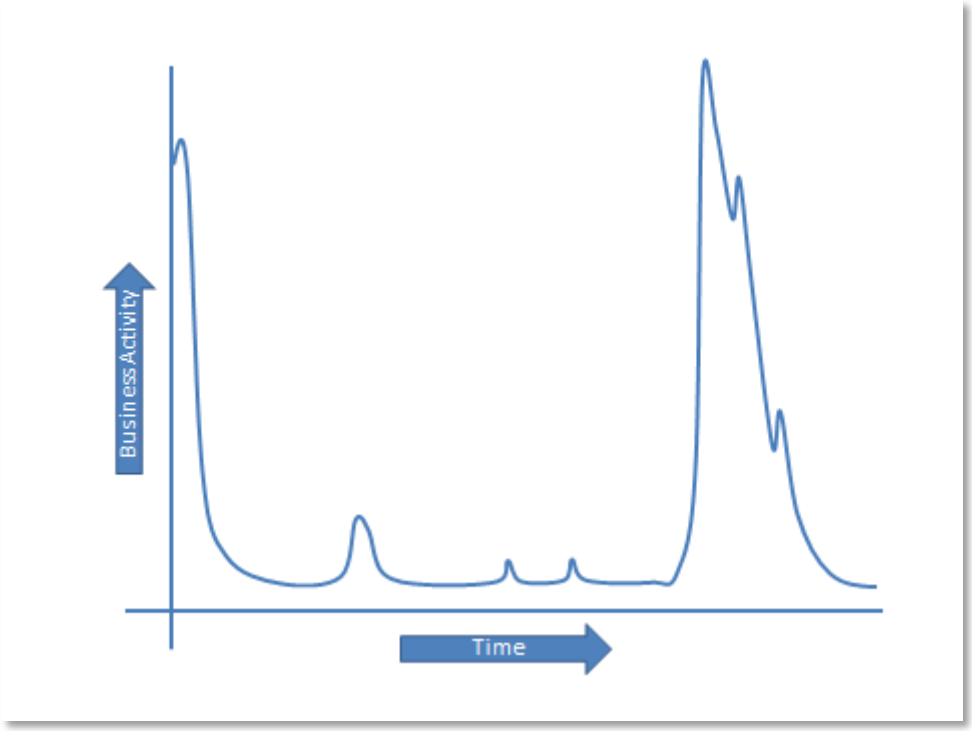


The public cloud offers firms the opportunity to tune their resource requirements minute-by-minute, so that when something unpredictable happens at a normally predictable moment, they can fully exploit the opportunities. The double-peak above allows extra business to be maximised. Conversely a double dip gives organisations the opportunity to minimise the resources they assign and thereby maintain the link between the cost of operations and the revenue/profit they are able to generate.

Unpredictable Bursting

Almost all businesses experience this problem. Some event, usually external to the organisation, triggers an unpredicted burst of business activity. It could be large, such as a TV celebrity giving a product endorsement to a small business's product on peak-time TV. It could be small, such as an article appearing in a niche publication which generates a modicum of interest: big enough interest to need to gear the business up to react to it, but small enough to make the idea of responding to it

in the traditional non-cloud way worthless – the peak of interest will have disappeared by the time the organisation can scale.



A good way to think of these scenarios is to think of businesses a little like people moving upwards on a moving staircase (an escalator in UK parlance!). There are many opportunities that companies have traditionally been unable to exploit because by the time they react, the opportunity has passed. This can be represented as taking a backward step on the moving staircase. All the people on the moving staircase are getting to their destination, but the ones who take more backward steps than others will get there more slowly. If there are many opportunities that can't be reacted to, it's like one of the staircase passengers taking many small backward steps. Those organisations that are able to react quickly have fewer backward steps and win the race. They are more agile than their competitors and they are able to increase revenues, profits, satisfaction and so on. They can increase customer retention because they are always in a position to react, on a minute-by-minute basis to the sentiments and idiosyncrasies of their customers.

It's in the interest of all firms to react to large and sustained increases/decreases in business activities. It's the combined effects of all the little changes over a period of time, plus the speed the business can react to the large changes which separates the agile business from the lumbering dinosaur.

Summary

The cloud offers a chance for the modern business to be more agile than its competitors by allowing it to react quickly while still maintaining profitability. IDG's Market Pulse report "Cloud Computing in Europe: Transformation is in the Air" said that among the 600 companies surveyed in the U.S., Europe and APAC, business agility is seen as the greatest impetus to adopt cloud computing. In fact in the same report, many CIOs understood that enterprise class cloud computing would help them evolve to "IT as a revolutionary asset for the company".

Microsoft UK 2011

Document: Business Agility and the Cloud

Author: Steve Plank, Microsoft UK

Website: www.microsoft.com/cloud